Balancing incoming traffic over multiple links

Juha Väisänen
Helsinki University of Technology
Laboratory for Telecommunications software and Multimedia
javaisan@cc.hut.fi

Abstract

This paper introduces four different ways to overcome the problem of balancing incoming traffic over multiple links. The four ways are setting of the MED parameter, prepending of the AS_PATH parameter, usage of communities and defining of more specific routes. They are explored in own sections and compared in conclusion section. This paper is intended to readers with basic knowledge of Border Gateway Protocol (BGP) and routing.

1 Introduction

To overcome the criticality of having only one backbone link without any backup route, many operators lease more than one links. These links can be either to one service provider or to two or more different service providers. When dealing with multiple links with most likely different link statistics (e.g. capacity, throughput, usage rate) balancing both incoming and outgoing traffic becomes key factor for efficiency. Outgoing traffic is quite easily balanced in own routers but to influence how incoming traffic is balanced one has to try to configure parameters on external routers. Four ways to do this are explained deeper; Modifying of MED parameter of the route, Prepending the AS_PATH length, using of BGP communities and more specific route tactic.

2 Modifying the MED attribute

In this section, we explore the MULTI_EXIT_DISC (MED) attribute and discover how it can be used to balance incoming traffic. Usage of the MED in balancing is quite easy but very restricted. It should be used only in a situation of two neighboring Autonomous Systems. Sec. 2.1 emphasizes important points about MED, Sec. 2.2 gives examples of using MED and Sec. 2.3 shows an example of route oscillation with MED.

2.1 How MED works?

The MULTI_EXIT_DISC (MED) parameter is one of the attributes included in BGP protocol. MED is a 32-Bit interger value representing the non-transitive BGP metric. MED belongs to the set of BGP routing information parameters like NEXT_HOP, AS_PATH and LOCAL_PREF. MED parameter can be used to differentiate exit and entry points between two ASs. The MED parameter values are not advertised to other Autonomous Systems with the other route information. [4]

This leads to the fact that MED is applicable balancing actor only in the situation of two or more connections between two neighboring Autonomous Systems (ASs). Typically this happens with two or more connections with the same (Internet Service Provider (ISP).

In the normal case MED values are not used when comparing route information received from different AS. With use of *bgp always-compare-med* command this restriction is overruled and MED values are taken into account in every case. [7] This can though lead to loops and route oscillation.

2.2 Using MED to balance traffic

The most simple case where MED can be used to balance incoming traffic is in the situation where local site has two connections to one ISP in neighboring AS. These connections have different capacities and therefore equal balancing is not preferred. With adjusting MED parameter most of the incoming load can be directed to use the connection with better capacity. MED parameter has to be adjusted for both of the connections since *bgp bestpath med missing-as-worst* command can be in effect on the terminating router and therefore the connection with no MED seem to be the worse. [5]

2.3 Route oscillation with MED

The other example [4] demonstrates a worst case situation where usage of MED leads to route oscillation. These oscillations are typically divided into two types with different requirements. On the Type 1 looping there are two conditions that both need to be met.

1. a single level Route Reflection or AS Confederation design is used in the network
2. the network accepts the BGP MED attribute from two or more ASs for a single prefix and the MED values are unique.

On the Type 2 oscillation there are also two conditions that need to be both met.

1. More than one tier of Route Reflection or Sub-ASs is used in the network
2. the network accepts the BGP MED attribute from two or more ASs for a single prefix and the MED values are unique.

Following example in figure 1 demonstrates a situation where Type 1 oscillation occurs with AS Confederation.

Ra has in the beginning BGP table 1 where route "10 100" is selected as best and advertised to Rd. Rd has after Ra's advertisement updated routing table 2 where route through Ra is selected as best because of lower IGP metric. Rd sends UPDATE/withdraw to Ra for "6 100" route it had advertised. Ra updates it's table 3. Former best route was not removed so previous best route is preserved. Periodic BGP scanner realizes that "6 100" route is better (lower IGP metric) and updates own table 4. Then Ra sends UPDATE to Rd with this new best route advertised. Rd receives Ra's UPDATE and modifies the routing table 5. Now Rd has "6 100, 0, 6" route as best because of lower MED value. This route is now advertised to Ra. Ra gets this UPDATE and modifies it's routing table 6. Now the table looks like table 1 and this shows the type 1 oscillation with AS confederation.

AS PATH	MED	NEXT HOP IGP Cost
*10 100	10	3
(65001) 6 100	0	7
6 100	1	2

Table 1: Ra BGP routing table

AS PATH	MED	NEXT HOP IGP Cost
6 100	0	6
*(65001) 10 100	10	4

Table 2: Rd BGP routing table

AS PATH	MED	NEXT HOP IGP Cost
6 100	1	2
*10 100	10	3

Table 3: Ra BGP routing table

AS PATH	MED	NEXT HOP IGP Cost
*6 100	1	2
10 100	10	3

Table 4: Ra BGP routing table

AS PATH	MED	NEXT HOP IGP Cost
*6 100	0	6
(65000) 6 100	1	3

Table 5: Ra BGP routing table

AS PATH	MED	NEXT HOP IGP Cost
*10 100	10	3
(65001) 6 100	0	7
6 100	1	2

Table 6: Ra BGP routing table

3 Prepending the AS PATH

In this section another way of balancing incoming traffic is introduced. Path prepending is probably the most effective way to balance incoming traffic. It can be even too effective sometimes. Sec. 3.1 emphasizes important points about PATH prepending and Sec. 3.2 gives example of prepending AS PATH attribute as a balancing act.

3.1 Principles of path prepending

AS PATH is a mandatory attribute specified in BGP protocol specification. It contains a sequence of segment triples <path segment type, path segment length, path segment value> defining either ordered or unordered set of ASs that the UPDATE message has traversed. [3]

In PATH prepending own AS number is appended to the path one or more extra times to make the length of the path look longer for neighboring ASs. Prepending AS PATH is very effective since it is updated over ASs and BGP protocol uses path length early when calculating best route. This can sometimes lead to even too effective change in traffic flow. [5]

This technique suits ideally in situation of end-user network but not so well for ISPs. This is because this new path length information is modified to external BGP peers and there to other ASs. Therefore in worst case this modification would lead to situation where no traffic flows through that ISP. [7]

3.2 How it really works?

This section is divided into two example situations originally covered in [5]. In the first case in the figure 2 multihome network is connected to two similar ISPs which then are connected to similar networks. In the second case in the figure 3 the network topology is more complex and the ISPs aren't anymore similar.

Here AS1 is connected to two ISPs. They are similar but AS1 wants to change the traffic flow and prepends the other path. In this situation there are three choices for traffic flow. Without path prepending traffic may or may not be balanced. When prepending path to ISP A, majority of the traffic flows through ISP B and when prepending path to ISP B, most of the traffic comes in over ISP A.

Action	Traffic Distribution
Prepend A	ISP A: 15%, ISP B: 85%
No Prepending	ISP A: 40%, ISP B: 60%
Prepend B	ISP A: 90%, ISP B: 10%

Table 7: Simple network prepending

In this example figures in the table 7 are based on assumptions that ISP A handles always 15% of the traffic, ISP B always 10% and ASs from 2 to 4 25% each. When no prepending is used ASs 2 and 4 prefer to send over ISP B and AS 3 over ISP A.

In this second example case multihome AS is connected to two dissimilar ISPs. These ISPs are dissimilar since they connect directly to different set of ASs and then buy transit service to other. So in this situation prepending path has more diverse effect.

Figures in the table 8 of the complex network are based on assumptions that ISP A handles always 15% of the traffic, ISP B and ASs 6 and 7 5%, 10% from AS2 and 20% from

Action	Traffic Distribution
Prepend A * 2	ISP A: 15%, ISP B: 85%
Prepend A	ISP A: 35%, ISP B: 65%
No Prepending	ISP A: 55%, ISP B: 45%
Prepend B	ISP A: 75%, ISP B: 25%

Table 8: Complex network prepending

ASs 3–5. Even when path to ISP B is prepended ASs 2, 6 and 7 use route through ISP B. When no prepending is used AS 5 prefers to send over ISP B.

4 Using of communities

Small introduction to the world of AS communities. This section introduces yet another way of balancing incoming traffic. Section 4.1 shows technical details behind communities and explains how they work. In Sec. 4.2 example of using communities is given.

4.1 What are communities?

BGP community is an optional variable size transitive attribute. It has variable amount of four octet sets each defining a community. First two octets have community attribute values coded with AS number and the other two octets are freely modified by the AS. A community is a group of destinations sharing a common property. RFC 1997 [1] introduces also three well-known global communities. These communities are NO_EXPORT (Routes received with this community value are not advertised to eBGP peers), NO_ADVERTISE (Routes with this community value are not advertised to any peer) and NO_EXPORT_SUBCONFED (Routes with this community value are advertised only to iBGP peer within same AS confederation.

These well-known communities are useful for example in a situation where ISP doesn't want it's customer to advertise it's route to other ISP. Proper caution should be used when using communities since misconfigured community attribute might lead to a situation where no routes are advertised even if some of them would be wanted.[5]

4.2 Using Communities

In many cases the set of accepted communities are listed in AUT_NUM object in the Routing Register. It's worth checking out this object before using a custom community value to avoid unwanted results. Normal communities that are used in addition to those well-known communities have actions like setting local preference value, prepending path to an ISP or an interconnect point and orders for announcing route for transit routers.

Communities are a useful way to handle the same functionalities as MED handles but even better since they can be used to set Local Perference values for connections. When Local Preference value is set for one connection to a sufficiently low level this connection becomes a backup connection that is used only when the main connection isn't available. Using this community frees MED parameter for other use and what's even better, is that these adjustments can be advertised also beyond AS border which is something MED wasn't capable.

In situation where ISP has many connections it is easier to handle their parameters with communities. ISP can use community to prepend AS PATH value for one connection and another community to prepend other connection. Power of communities lies in the fact that many community values can be easily added for one connection and that by using communities same functionalities can be handled as with MED and AS PATH prepending.

5 Specific Routes

This section explains setting of more specific routes as a balancer for incoming traffic. This is the final thing to do if nothing else seems to work. This procedure changes the global routing tables so it should be done only when extremely necessary. [8]

5.1 What are specific routes

In this technique some routes are given more specific attributes. Routes with more specific information are selected over less specific ones. Therefore this technique is guaranteed to work as long as advertised ISP and other parties in the upstream way accept these extra specifics. For example in the situation of figure 2 if routers for ISP A use a lower Router ID than routers of ISP B then most of the traffic would flow over ISP A because they got higher IP addresses. When a tie-break situation occurs for two equally long AS paths, the route with lower router ID is selected.

This more specific route tactic can be explained with a situation where AS 1 from figure 2 has address block of 192.168.0.0/20 (16 Class C addresses: 192.168.0 through 192.169.15). AS 1 then announces address block 192.168.0.0/21 for ISP A and rest of the block for ISP B. Then all traffic coming to addresses in block 192.168.0.0/21 come through ISP A and traffic to addresses 192.168.8.0/21 over ISP B.

When using this address block sharing tactic it should be payed attention that high bandwith host are distributed equally in the case of two similar connections or to use high bandwith connection in case of dissimilar connections.[8]

6 Conclusion

In this paper four ways to balance incoming traffic were introduced. MED was easy to implement but had restricted functionalities . AS_PATH prepending was efficient but harder to implement and needed better knowing of the network topology. Communities were sort of combination of everything. They were used to handle many functionalities like path prepending and how routes were advertised. Using of specific routes was definitely the most troublesome to implement and very accurate information about network was needed. All of the introduced tactics have their pros and cons. MED is suitable for small networks because of it's restrictions and easiness. PATH prepending is most suitable for situations where MED isn't enough and more power is needed. Communities are the best choice when number of connections is large and many connections would have the same parameters. Then setting up communities to handle route advertising and traffic balancing is efficient. Setting of more specific routes is good and working practice but still it should not be used whenever the balancing can be handled otherwise.

References

[1] R. Chandra, P. Traina and T. Li. BGP Communities Attribute. RFC 1997, IETF Network Working Group, August 1996.

[2] E.Chen and T. Bates. An Application of the BGP Community Attribute in Multi-home Routing. RFC 1998, IETF Network Working Group, August 1996.

[3] P. Traina, D. McPherson and J. Scudder. Autonomous System Confederations for BGP. RFC 3065, IETF Network Working Group, February 2001.

[4] D. McPherson, V. Gill, D. Walton and A. Retana. Border Gateway Protocol (BGP) Persistent Route Oscillation Condition. RFC 3345, IETF Network Working Group, August 2002.

[5] Iljitsch van Beijnum. Traffic Engineering: Incoming Traffic. O'Reilly, OnLamp.com, www.onlamp.com/lpt/a/2797, 2002

[6] Peter Morrissey. Multihoming with BGP4. Network Computing, www.networkcomputing.com/1025/1025ws1.html , December 1999.

[7] Iljitsch van Beijnum. Traffic Engineering: Local Routing Policy. O'Reilly, OnLamp.com, www.onlamp.com/lpt/a/2796, 2002

[8] Iljitsch van Beijnum. Traffic Engineering: Specific Routes. O'Reilly, OnLamp.com, www.onlamp.com/lpt/a/2798, 2002

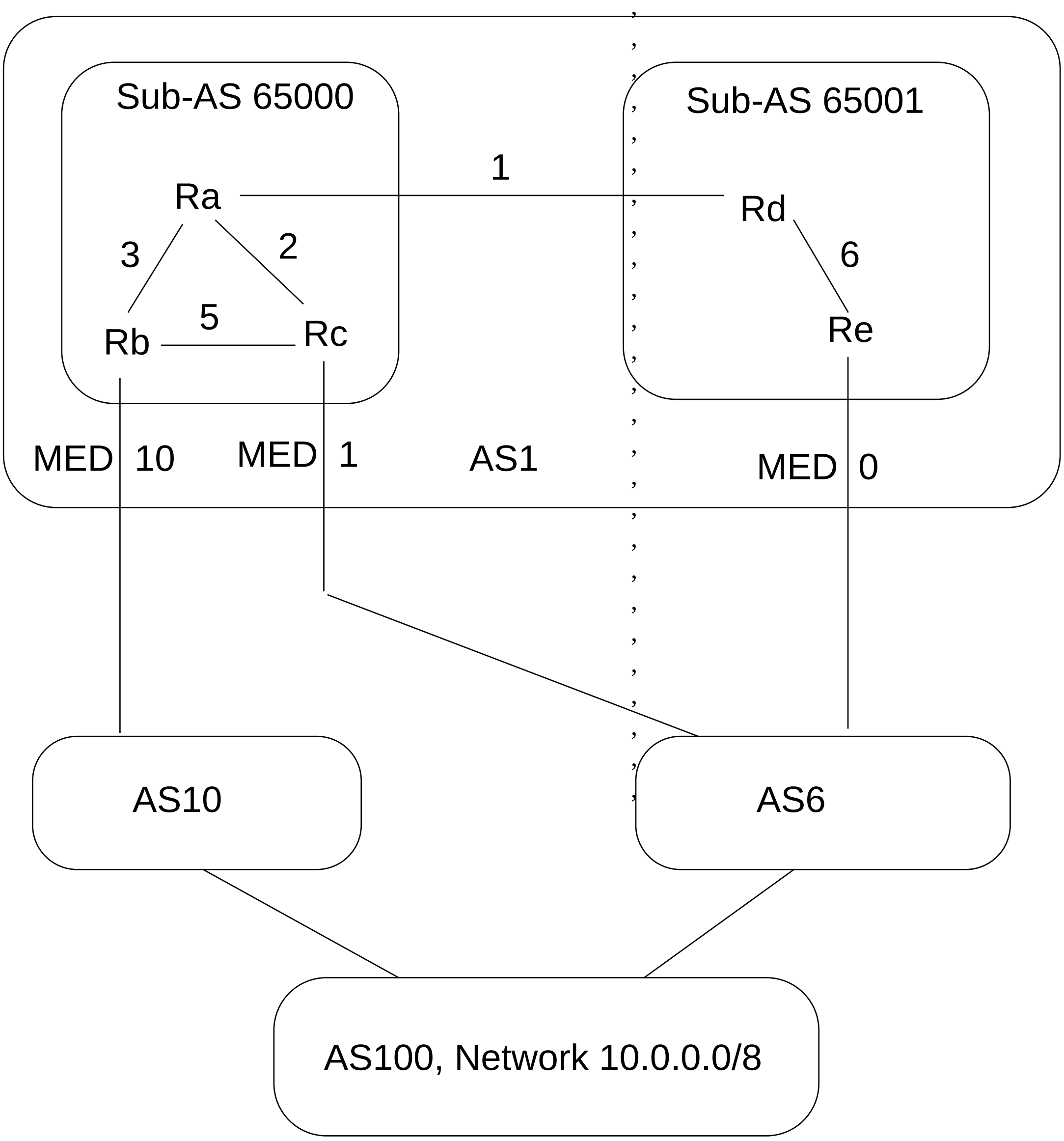

Figure 1: Example MED Network

,
,
,
,
,
,
,
,
,
,
,
,
,
,
,
,
,
,

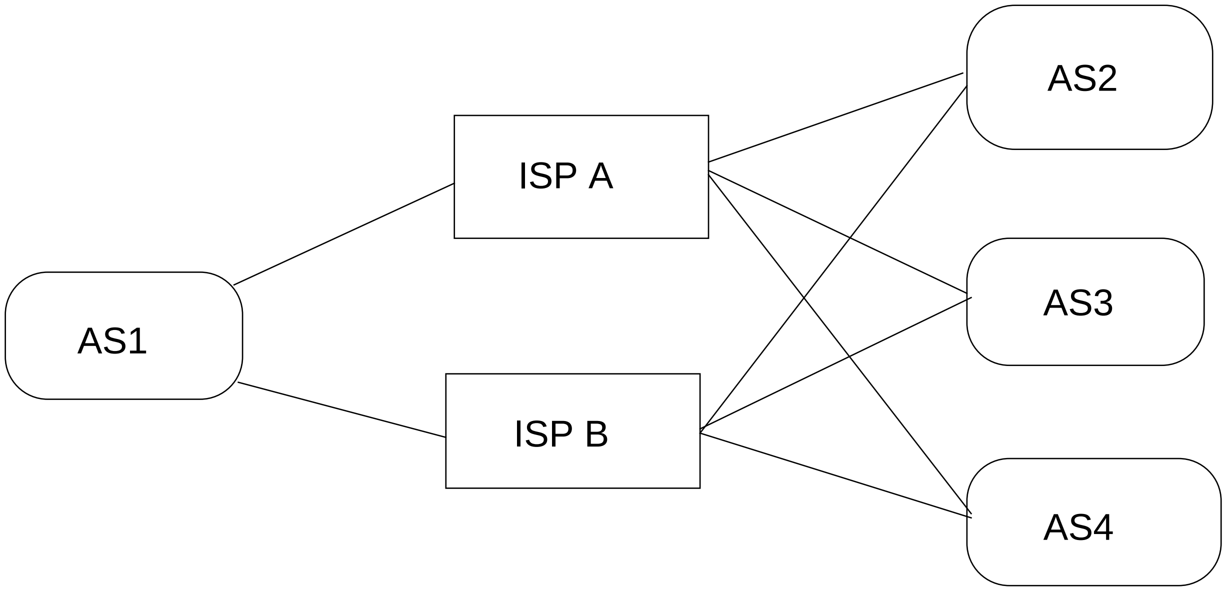

Figure 2: Example Network 1

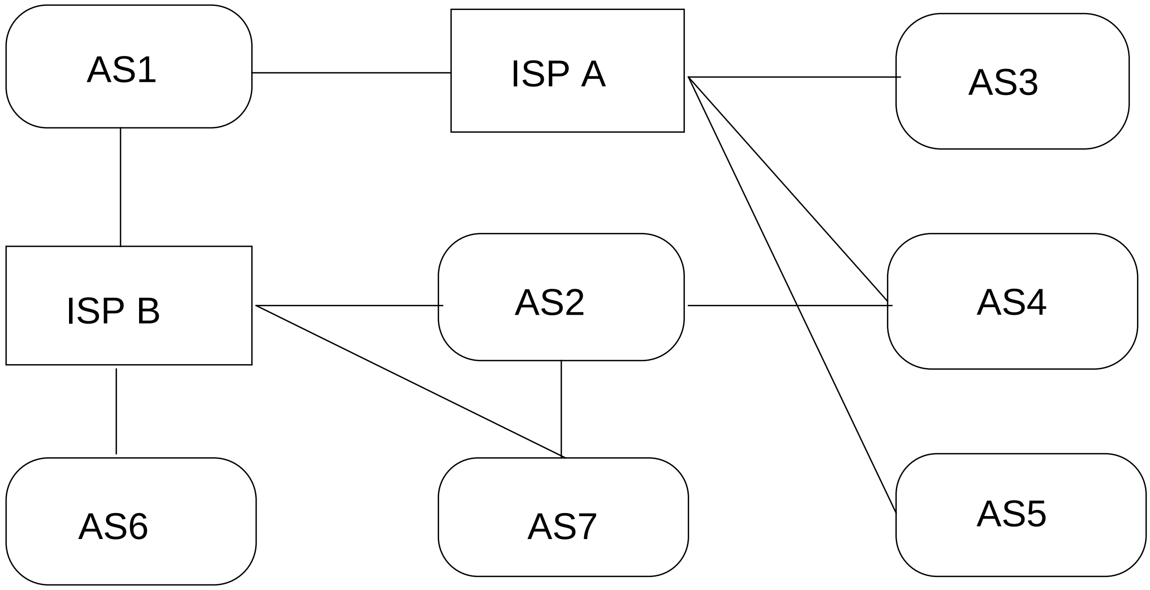

Figure 3: Example Network 2